

# 가변계수 프리엠퍼시스를 이용한 마이크어레이 음성입력에 관한 연구

조왕래, 배명진

승실대학교 정보통신공학과 음성통신연구실

## On Speech Input with Microphone Array using the variable coefficient Pre-emphasis

Wangrae Jo, Myungjin Bae

Dept. of Telecommunication Engr., Soongsil University

### 요 약

현재 음성인식기는 다채널의 음성입력방식을 사용하고 있는 추세이다. 이런 방법으로 음성인식기를 사용할 때에 자동적으로 음성을 검출하는 음성입력방식은 발성자와 마이크 간의 거리에 따라 발성음성, 반사음성과 잡음이 입력된 경우에 원 음성의 고차포만트 성분에 왜곡이 발생하게 된다. 이러한 문제점을 개선하기 위하여 본 논문에서는 고주파영역의 특성을 강조시킬 수 있는 프리엠퍼시스를 적용하여 더욱 더 정확한 음성입력방법을 제안한다. 제안한 방법은 원음과 유사한 합성입력음을 얻었고, 또한 기존의 시간영역법보다 더 높은 SNR을 얻을 수 있다는 것을 알 수가 있다.

### 1. 서론

최근 십여년간 원거리 화상회의, 녹음실에서의 녹음, 음성인식을 위해서 마이크로폰 어레이를 사용하는 방법에 대한 많은 연구가 이루어지고 있다. 현재 발표되는 대부분의 음성인식시스템은 하나의 마이크를 이용하여 음성을 입력하는 방법을 채택하고 있다. 그러나, 하나의 마이크로 음성을 입력받는 경

우에는 음성 입력시에 마이크의 위치에 항상 세심한 주의를 기울여야 한다. 특히 음성 인식시스템은 높은 신호대잡음비를 얻어야 하기 때문에 마이크를 가까이에 두고 발성하여야만 원하는 수준이상의 인식률을 얻게 된다. 이러한 경우 일반인들이 실생활에 사용하기에는 여러 가지로 불편하다. 본 논문은 사용자가 마이크의 위치에 주의를 기울이지 않고 발성하는 경우에도 허용된 수준이상의 신호대잡음비를 갖는 음성을 자동으로 입력해 주도록 하기 위해 다채널 마이크 어레이를 이용하는 방법에 관한 것이다.

### 2. 기하학적 모델

그림 2-1은 2차원 기하학 모델과 M개의 센서들을 가지고 있는 선형 마이크로폰 어레이를 위한 일반적인 음원모델을 보여주고 있다. 마이크로폰어레이들은 그들의 위치에 따라  $P_0(x_0, y_0), P_1(x_1, y_1), \dots, P_{M-1}(x_{M-1}, y_{M-1})$ 과 같이 표현될 수 있다. 음원에서는 음원신호  $x(t)$ 를 발생하고, 마이크로폰 0, 1, ..., M-1에 신호  $s_0(t), s_1(t), \dots, s_{M-1}(t)$ 가 입력된다. 각 음원신호  $x(t)$ 에 대해 t 번째 마이크로폰에 의해 입력된 신호는 다음과 같이

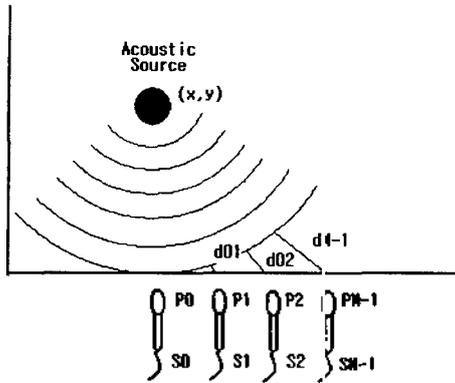


그림 2-1. 선형마이크로폰 어레이

표현될 수 있다.

$$s_i(t) = a_i r(t - \tau_i) + n_i(t) \quad (2.1)$$

여기서  $a_i$ 는 전파효과로 인한 감쇄상수이고,  $\tau_i$ 는 전파시간,  $n_i(t)$ 는 음원신호  $r(t)$ 와 독립인 잡음을 의미한다.  $i$ 번째 마이크로폰과  $j$ 번째 마이크로폰 사이의 상대적인 지연값  $\delta_{ij}$ 를 다음과 같이 나타낼 수 있다.

$$\delta_{ij} = (\tau_i - \tau_j) \quad (2.2)$$

따라서 다채널 마이크로폰 어레이들 사이에서 상대적인 지연값들을 정확히 추정한다면 원음정보보다 신호대잡음비가 높은 합성음을 얻을 수 있다.

### 3. 시간 영역 다채널 마이크 어레이 음성 입력

상대적인 지연정보를 정확하게 추정하기 위한 전처리 단계로서 음성구간을 검출한다. 본 논문에서는 가장 보편적인 방법으로 사용되고 있는 단구간 에너지와 영교차율을 이용하는 방법을 사용하여 음성구간을 검출한다.

음성검출이 종료된 후에는 검출된 음성부

를 이용하여 채널별 음성간의 시간지연 정보를 추정하고 시간적으로 동기가 일치하게 처리한 후에 다채널의 음성을 가산하여 잡음이 감소된 입력음성을 얻는다. 이 과정에서 잡음감소 효과를 높이기 위해서는 각 채널의 음성신호간의 지연정보를 정확히 추정해야 한다. 본 논문에서는 각 채널의 상호상관 관계를 이용하여 시간 지연정보를 추정하였다. 즉, 기준채널의 음성신호와 나머지 한 채널의 음성신호간의 상호상관 함수를 구한 다음, 상호상관 함수가 최대가 되는 지점을 기준 채널의 음성신호와 비교채널의 음성신호간의 지연 시간으로 정하였는데 식(3.1)과 식(3.2)와 같이 나타낼 수 있다.

$$R_{ij} = E[s_i(t)s_j(t + \tau)] \\ = a_i a_j R_{rr}(\tau - \tau_{ij}) + R_{n,n}(\tau), \quad (3.1)$$

$$i = 0, j = 1, \dots, M-1$$

$$\tau_{ij} = \operatorname{argmax} R(\tau) \quad (3.2)$$

$$= \operatorname{argmax} \sum_{n=0}^N s_i(n)s_j(n + \tau)$$

여기서  $m$ 은 마이크어레이의 개수를 나타내고,  $s_i$ 와  $s_j$ 는 각각 기준 채널의 음성신호와 비교 채널의 음성신호를 나타낸다.

추정된 각 채널별 음성의 시간지연 정보로부터 이를 상쇄시키는 과정을 거친 다음, 각 채널의 잡음 신호의 에너지 크기가 같아지도록 정규화하고 최종적으로 시간적으로 동기가 일치된 음성신호를 서로 가산하여 식(3.3)과 같이 다채널 신호 처리된 음성부를 얻는다.

$$\tilde{s}(n) = \frac{1}{M} \sum_{k=0}^{M-1} s_k(n + \tau_k) \quad (3.3)$$

그러나 이와같이 시간영역에서 동기가 맞춰서 각각의 채널 신호를 가산하였기 때문에 각각의 마이크에 입력될 때 잡음이 섞이는

경우에 포맷트 정보, 특히 고차 포맷트 정보에 왜곡이 발생하기 때문에 고차포맷트의 특징성분에 둔감해지게 된다. 이러한 문제점으로 정확한 동기정보를 찾을 수 없다는 단점이 있다.

#### 4. 가변계수 프리엠퍼시스를 이용한 마이크 어레이 음성 입력

음성신호  $s(n)$ 을 주파수에서 평탄화하기 위해 그리고 신호처리에서 유한 정밀도 등의 효과에 덜 민감해지도록 저차시스템에 통과되어야 한다. 가장 널리 사용되는 시스템은 다음과 같은 고정된 1차 시스템이다.

$$H(z) = 1 - az^{-1}, \quad 0.9 \leq a \leq 1 \quad (4.1)$$

이 경우에 프리엠퍼시스의 출력은 다음과 같은 차분방정식으로 표현되어 질 수 있다.

$$\hat{s}(n) = s(n) - as(n-1) \quad (4.2)$$

여기서  $a$ 값은 보통 0.9~1사이의 값을 주로 사용한다.

음성신호를 관찰해보면 성문특성으로 인해 고주파 특성이 약화되는 것을 알 수 있다. 이러한 현상은 마이크로부터 거리가 멀어지면 멀어질수록 심해진다. 이러한 감쇠기울기를 측정하여  $a$ 값을 달리 조정하여 프리엠퍼시스를 적용하면 좀 더 정확한 지연값을 추정할 수 있다. 본 논문은 각 마이크 입력 신호의 고차포맷트의 성분을 강조하기 위해 기울기에 따른 가변계수 프리엠퍼시스 필터를 통과한 뒤 더욱 세밀한 지연값의 추정함으로써 기존의 방법보다 더 높은 신호대잡음비를 얻을 수 있는 방법이다.

#### 5. 실험 및 결과

실험에 사용된 음성데이터는 PC를 사용하고 있는 일반적인 실험실 환경에서 수집한 단어중에서 2명의 화자가 발성한 문장을 사

용하였다. 이들 데이터는 11kHz로 표본화되고 16bit로 양자화 되었으며 1프레임은 256샘플로 처리하였다.

마이크 어레이를 구성하는 마이크들은 발화자 전면의 모니터의 네 모퉁이에 설치하였다. 좌측과 우측 마이크사이의 거리는 42cm이었으며, 상하단 마이크사이의 거리는 35cm 정도이었다. 발화자는 모니터로부터 40cm 떨어진 위치에서 세 번씩 발성하였다.

각 채널의 입력 음성을 식(3.1)과 식(3.2)와 같은 방법으로 시점을 동일하게 맞춘 다음 가변계수 프리엠퍼시스에 통과시켜 정확한 지연정보를 추출하는 방법으로 마이크 어레이 출력음성을 얻었다.

성능평가를 위해 화자 1, 2가 중앙, 좌측, 우측에서 발성한 음성을 기존의 방법과 제안한 방법으로 추출한 출력신호의 SNR을 비교하여 표 5-1과 표5-2에 나타내었다. 결과를 살펴보면 제안한 방법이 기존의 방법보다 4.69~8.13dB 정도의 향상된 음성출력을 얻을 수 있었다.

표5-1. 기존의 방법과 제안한 방법의 SNRseg(화자1)

	발성위치		
	좌측	중앙	우측
기존의 방법	7.22	6.29	7.54
제안한 방법	13.88	14.42	13.95
편 차	6.66	8.13	6.41

표5-2. 기존의 방법과 제안한 방법의 SNRseg(화자2)

	발성위치		
	좌측	중앙	우측
기존의 방법	7.06	8.31	6.70
제안한 방법	13.02	13.00	12.59
편 차	5.96	4.69	5.89

## 6. 결론

마이크의 위치에 구애받지 않고 음성을 효과적으로 입력하기 위한 한 방법으로 4개의 마이크어레이를 사용한 4채널 음성입력방법을 제안하였다. 이는 4개의 마이크에서 얻어진 각각의 음성신호에 가변지수 프리엠퍼시스를 적용함으로써 교차포먼트의 왜곡에 관계없이 더욱 정확한 동기정보를 얻을 수 있으며 이를 적용하면 마이크의 위치에 구애받지 않고 더욱 정확한 입력음성을 얻을 수 있다. 실험 결과 제안한 방법은 기존의 방법에 비해 약 4.69~8.13dB정도의 신호대잡음비가 향상되었다. 앞으로 제안한 방법을 자동 음성인식시스템에 적용하였을 때의 인식율 향상에 관한 연구가 이루어져야 하겠다.

## 참고문헌

1. A.M.Kondoz, "Digital Speech", John Wiley & Sons Ltd., 1994.
2. G. Clifford Carter, "Coherence and Time Delay Estimation", Proceedings of the IEEE, vol. 75, no. 2, pp.:236-255, Feb. 1987.
3. 이순민, "성문특성이 제거된 성도특성 추출에 관한 연구", 숭실대학교 석사학위논문, 2001.
4. 서영주, 이항섭, 이영직, "마이크어레이와 자동 음성부 검출을 이용한 원격음성입력기의 구현", 한국음향학회 학술발표대회 논문집, vol.15, 한국음향학회, 1996.
5. Maurizio Omologo, "Acoustic Event Localization Using a Crosspower-Spectrum Phase based Technique", Proc. ICASSP94 of the IEEE., 1994.